Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB05/001218

International filing date: 24 March 2005 (24.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: GB
                         Number: 0406800.3
                         Filing date: 26 March 2004 (26.03.2004)

Date of receipt at the International Bureau: 24 May 2005 (24.05.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

GB /05 / 1218

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.
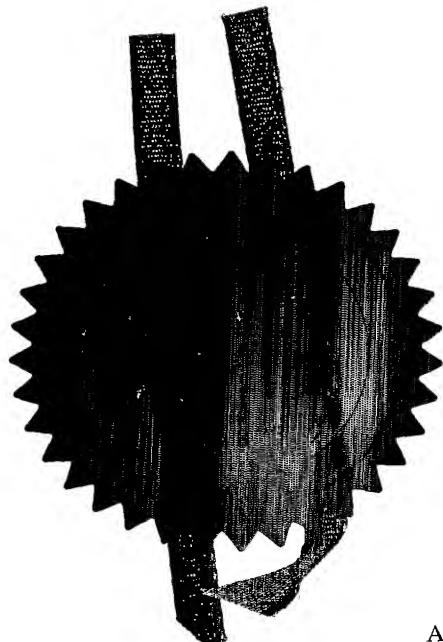
Signed

Dated   12 April 2005
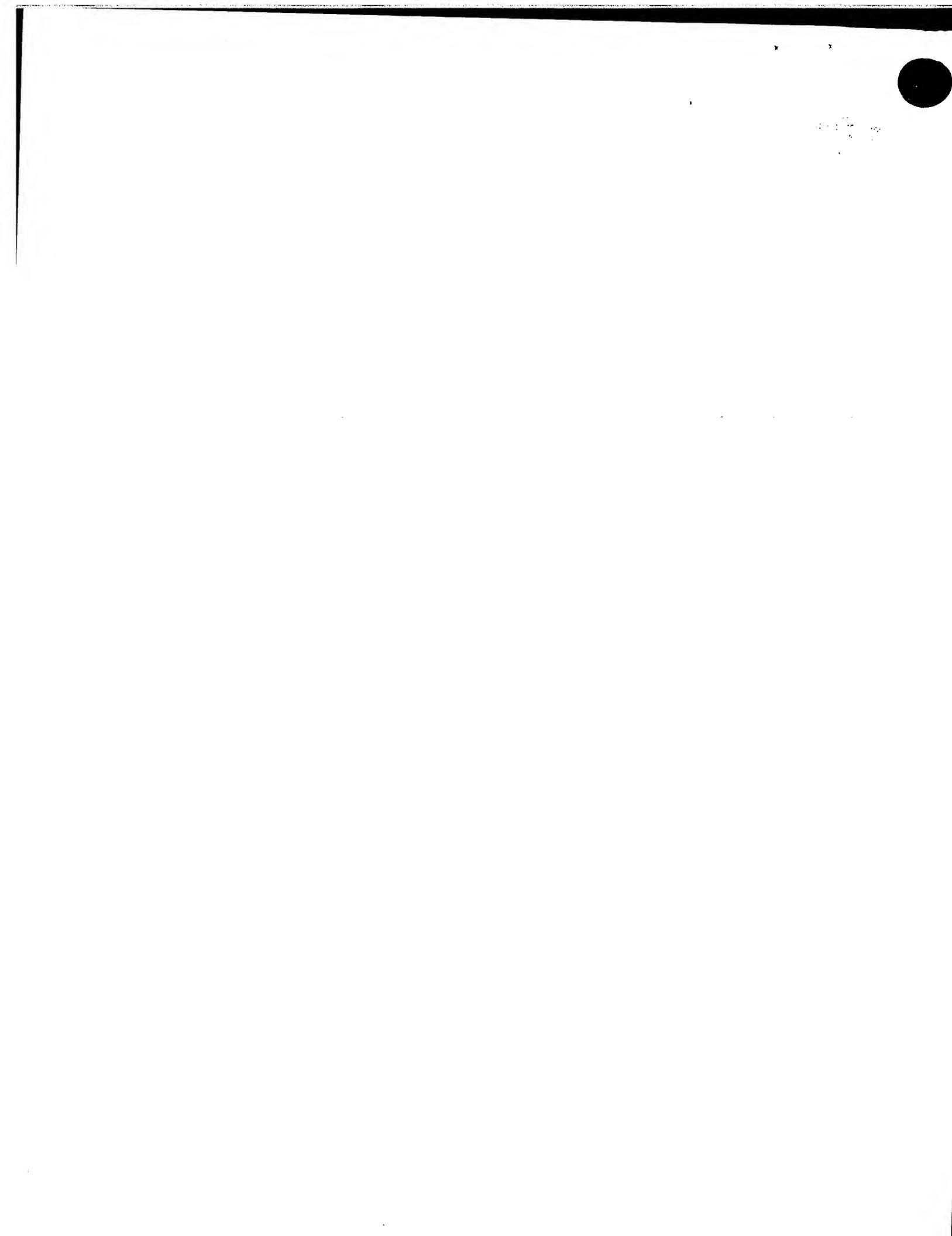
The Patent Office

26MAR04 E884132-1 811571
P01/7700 0.00-0406800.3 NONE

2 6 MAR 2004

Request for grant of a patent

*(See the notes on the back of this form. You can also get
an explanatory leaflet from the Patent Office to help you fill in
this form)*

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1.	Your reference	MICROARRAY METHOD

2.	Patent application number *(The Patent Office will fill this part in)*	**0406800.3**	2 6 MAR 2004

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

THE UNIVERSITY OF NOTTINGHAM
UNIVERSITY PARK
NOTTINGHAM
NG7 2RD

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation

4376927002

4. Title of the invention

MICROARRAY METHOD

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

N/A

Patents ADP number *(if you know it)*

		Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*
6.	Priority: Complete this section if you are declaring priority from one or more earlier patent applications, filed in the last 12 months.	N/A		

			Number of earlier UK application	Date of filing *(day / month / year)*
7.	Divisionals, etc: Complete this section only if this application is a divisional application or resulted from an entitlement dispute *(see note f)*	N/A		

8. Is a Patents Form 7/77 (Statement of inventorship and of right to grant of a patent) required in support of this request?

Answer YES if:
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
Otherwise answer NO (See note d)

YES

Patents Form 1/77

9. Accompanying documents: A patent application must include a description of the invention. Not counting duplicates, please enter the number of pages of each item accompanying this form:

Continuation sheets of this form

Description	7
Claim(s)	—
Abstract	1
Drawing(s)	—

10. If you are also filing any of the following, state how many against each item.

Priority documents	—
Translations of priority documents	—
Statement of inventorship and right to grant of a patent (Patents Form 7/77)	1
Request for a preliminary examination and search (Patents Form 9/77)	—
Request for a substantive examination (Patents Form 10/77)	—
Any other documents (please specify)	—

11. I/We request the grant of a patent on the basis of this application.

Signature(s) *Shuxbarbh*   Date 24-MArch-2004

12. Name, daytime telephone number and e-mail address, if any, of person to contact in the United Kingdom

JANE OSMENT   0115 9515673

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered YES in part 8, a Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) Part 7 should only be completed when a divisional application is being made under section 15(4), or when an application is being made under section 8(3), 12(6) or 37(4) following an entitlement dispute. By completing part 7 you are requesting that this application takes the same filing date as an earlier UK application. If you want the new application to have the same priority date(s) as the earlier UK application, you should also complete part 6 with the priority details.

Aug 03

Patents Form 1/77

Microarray Method

Field of Invention

This invention relates to a technique for analysing cross-species microarray data.

Background of the Invention

High-density oligonucleotide microarrays (see Patent numbers WO9710365, EP0853679, US6040138, DE69625920T) are commonly used to study global gene expression in several organisms for which complete or extensive genome sequence information is available. In agriculture such high throughput approaches could be used to select more environmentally friendly chemicals for plant protection and to develop plants with increased yield, better nutritional value and increased resistance to diseases. The technology could be used in any analysis where cross-species hybridisation is possible including but not limited to prokaryotic, mammalian, plant nucleic acid sequences. However, lack of genomic sequence information for most crop species is a major limitation to gene expression profiling with high-density oligonucleotide arrays.

A tailored approach to cross-species hybridisation of crop species nucleic acid high-density oligonucleotide arrays designed for model or extensively sequences organisms such as Arabidopsis is a novel solution. For the purposes of this application cross-species hybridisation is taken to mean the hybridisation of nucleic acid fragments of one species with nucleic acid fragments of a related species.

The high-density oligonucleotide microarray platform uses between 11 and 20 pairs of oligonucleotide probes, called probe sets, to provide the complete sequence of the gene (the 'target'). The array platform comprises perfect match probes and mismatch probes. The perfect match probes represent identical copies of the target sequence and the mismatch probe sequences differ from the perfect match sequences at the central nucleotide position. The mismatch probes measure non-specific binding. These oligonucleotide DNA probes are tethered by covalent attachment to a solid support to form an array platform (for example, the Affymetrix GeneChip®).

Typically, during gene expression profiling, the mRNA target, which has been hybridized to probes on the array, is fluorescently labelled. The hybridisation signal, computed as the difference between the perfect match and the mismatch probes across a probe set represents a measure (also referred to as the expression estimate) of gene expression level. The hybridisation of target mRNA transcripts from a species other than from the species (here referred to as the extensively sequenced or 'reference species') for which the oligonucleotide probes tethered on the array were selected, is termed cross-species hybridisation.

Prior to this invention several cross-species techniques (Chismar et al. Biotechniques 33 (3); 516-524 (2002): Caceres et al. Proc Natl Acad Sci USA, 100 (22); 13030-13035, (2003): Higgins et al. Toxicological Sciences, 74 (2); 470-484 (2003): Becher et al. The Plant J, 37; 251-268 (2004): Weber et al. The Plant J, 37; 269-281 (2004)) have been proposed, however, none of these techniques are in routine use. Most of these techniques were designed specific for the species under investigation and therefore inaccessible to most scientists. Most importantly, this prior art fails to adequately deal with the integrity of expression estimates.

Since high-density oligonucleotide arrays use 11-20 probes to interrogate each transcript it is conceivable that some of the probe sequences will differ from the cross-species sequence (i.e. the mRNA target). This poor sequence identity leads to inefficient hybridisation between array probes and the cross-species target. These probes, with poor sequence identity, generate weak hybridisation intensities which lead to an attenuation of hybridisation signals of probe sets. Computation of expression estimates for the cross-species target using all probes (11-20) of a probe set, will therefore generate inaccurate expression estimates.

Prior to the present invention, it has been proposed by others (Zhu et al. J Assoc. Lab. Automat, 6: 95-98 (2001)) that for cross-species microarray analysis, it is essential to pre-select usable probes by initially hybridising genomic DNA to the high-density oligonucleotide array. However, the prior art did not provide a rapid and effective method for identifying usable probes for cross-species data analysis. Probe selection by genomic DNA hybridisation prior to and independent of target mRNA hybridisation to the array, provides the most effective means of identifying usable probes as explained below.

Description of the Invention

This invention provides for a method of selecting a subset of probes, represented on a high-density oligonucleotide array, for use in the analysis of cross-species data, generated by hybridising the cross-species mRNA transcripts to the 'reference species' array.

In a preferred embodiment, oligonucleotide probes can be selected by hybridising genomic DNA or RNA to the array. Methods of isolating genomic DNA are well known to those skilled in the art. The DNA can be labelled for hybridisation by using the Bioprime ® DNA labelling System (Invitrogen). Instructions for use are provided by the manufacturer. Procedure for genomic DNA hybridisation follows that for target mRNA (i.e. antisense cRNA) hybridisation as fully described in the Patent Application DE69625920T but without the use of the cRNA hybridisation controls (Bio B, Bio C, Bio D and Cre). After staining and scanning the array as described in the Patent Application DE69625920T, the software which is described in detail for example U.S. Pat. Nos. 5,547,839, 5,578,832, 5,631,734, generates a hybridisation intensity file (CEL) containing statistics of each probe on the array, e.g., the 75

percentile of intensities, standard deviation of pixel intensities and probe co-ordinates which represent the physical location of the probes on the array.

A significant part of this invention is the construction of a series of programs using the scripting computer language PERL (see e.g., Wall, Christiansen, and Orwant, Programming Perl, 3rd Ed, O'Reilly and Associates (2000)) although equivalent scripts and programs can readily be developed in other computing languages by a person skilled in the art (including but not limited to C++, Java, Visual Basic). First, a perl script is constructed to extract probe co-ordinates with hybridisation intensity above background. The term background refers to non-specific hybridisation or other interactions between the hybridising target and components of the array. Fluorescence of the array components may also contribute to background. In one aspect of this preferred embodiment, background is calculated as the mean intensity of negative control probes. These negative control probes (Bio B, Bio C, Bio D, Cre etc) are oligonucleotide probes selected from species other than the 'reference species' organism or the cross-species organism.

The oligonucleotide probe co-ordinates selected by the perl script include both perfect match and mismatch probes. In a cross-species genomic DNA hybridisation it is conceivable that some of the mismatch probes will hybridise more efficiently to the cross-species target DNA sequence.

In a further aspect of this embodiment of the invention, a second perl script is developed to eliminate mismatch probes with hybridisation intensity above background and with higher hybridisation intensity than perfect match probes. To achieve this, the perl script uses as input, a file consisting of only perfect match co-ordinates and the output file of the first perl script. The output file generated in this embodiment consists of only perfect match probe co-ordinates with hybridisation intensities greater than the estimated background for the genomic DNA hybridisation. These perfect match oligonucleotide probes, which share high sequence similarity with the cross-species target, represent probes selected for the analysis of cross-species transcripts hybridised to the 'reference species' array.

The selection of the corresponding mismatch oligonucleotide probes is carried out as described in the third aspect of the invention.

In a third aspect of the preferred embodiment of the invention, a third perl script is constructed to complete the process of generating a chip description file (CDF) for the cross-species organism. This perl script uses as input, the output (selected perfect match probes) of the previous script and the chip description file of the 'reference species' organism. The X co-ordinate for both perfect match and mismatch probes for each gene sequence on the array is identical. This information is coded into the perl script to enable the selection of both perfect match and mismatch probes from the 'reference species' organism's CDF to construct a new chip description file for the cross-species. In this same aspect, the probes excluded from CDF construction can be used to construct a probe sensitivity index (PSI) file. The PSI file can be used to train a large data set in order to ensure consistency of data stored in a database.

All software used in the computation of gene expression levels require a hybridisation intensity file (CEL) of the type described above and a chip description file (CDF) for

the array type carrying the hybridised target transcripts. The chip description file is a library file consisting of gene (probe set) IDs, the corresponding co-ordinates of their probe sequences on the array and other software parameters.

In a second embodiment of the invention, a BLAST (Altschul et al., J. Mol. Biol. 215; 403-410 (1990)) output file is generated *in silico* by comparing cross-species nucleic acid sequences in a database to nucleic acid sequences represented on the oligonucleotide array. This output file is parsed with another perl script to identify oligonucleotide probes with 100% sequence identity to the cross-species sequence. The probes selected can then be used to construct a chip description file for the cross-species organism as described above.

The novelty of the invention in units of two embodiments lies in the informed selection of probe pairs (putative or actual match and mis-match) to allow high throughput genome-wide screening of the expression patterns of genes from genomes of species related to but not identical to the genome of at least one reference extensively-sequenced species such as but not limited to Arabidopsis, tomato, rice, mouse, Human, C. elegans, Bacillus sp., Drosophilia, Chimpanzee, Chicken, SARS virus etc.

This invention is a major advancement in microarray analysis of cross-species data. In particular for species where sequence information is not yet available. The reduction to practice of this invention means it is readily accessible to scientists investigating global gene expression in for example Brassica crops (oilseed rape, broccoli, cabbage, cauliflower, Brussels sprouts, radish, horseradish etc), but the invention is applicable to any species with significant synteny to an extensively sequenced species.

The invention is demonstrated in the following non-limiting examples.

Examples

Figure 1 shows gene expression levels computed with GeneChip® software (available from Affymetric Inc., Santa Clara, CA, USA).

There are three main columns, the first column (without title) consists of gene (probe set) IDs. The second and the third columns, which are subdivided into two columns denoting number of probe pairs (Stat Pairs) and expression levels (Signal), represent a sample (Bo+P) analysed with the model organism CDF (ATH1 – *Arabidopsis thaliana*) and the cross-species CDF (Brassica).

Notice the high level of expression values generated with the cross-species CDF. Since the mean of perfect match and mismatch probe pairs in a probe set is output by the software as the expression estimate of transcripts, probes in the ATH1 CDF which are not responsive to Brassica transcripts will lead to an attenuation of signal for the probe set, generating an inaccurate expression estimate of the transcript. In a typical experiment the background signal is usually less than 100. Therefore transcripts with expression levels below background are usually identified as undetectable in the sample being interrogated.

Figure 2 shows dChip (Li and Wong, Proc. Natl. Acad. Sci. USA, 98; 31-36 (2001)) software data of probe response patterns for the probe set, 246745_at.. The black graph (PM arrows) represents perfect match probe intensities and the grey graph (MM arrows) represents mismatch probe intensities. There are two grids, A and B, one depicting probe response pattern for 246745_at on an array (ATH1-12501) hybridized (see Patent No DE69625920T for all hybridization conditions) with an Arabidopsis (B1798_rep1) cRNA target (grid A) and a probe response pattern for 246745_at on the same array type hybridized with Brassica (Bo+P_rep1) cRNA target. There are 11 probe pairs in each grid, for grid A all eleven probe pairs respond actively to the Arabidopsis target. However, for the Brassica (cross-species) target only a subset (see grid B) of the 11 probes respond to the target. This is evident by the fact that for probe pairs 1-7 (grid B) the PM and MM curves are indistinguishable. Ymax for the graphs represents the highest probe signal intensity. These unresponsive probes were eliminated from the Arabidopsis (ATH1) chip description file to construct the cross-species (Brassica) chip description file.

Fig 1

Microarray Suite - [Expression Analysis]

File   Edit   View   Graph   Analysis   Run   Tools   Window   Help

Open   Save   Print   Tree   Shorts   Status   Expts   Fluidics   Scann

Find   Sort   Hide   Hide   Unhide   Options   Clear   Scatter   Series

	ATH1_Bo+P		Brassica_Bo+P	
	Stat Pairs	Signal	Stat Pairs	Signal
250555_at	11	1.4	2	300.5
263634_at	11	1.9	3	1091.4
260286_at	11	2.2	3	299.9
247304_at	11	2.8	4	305.7
259177_at	11	3.1	3	338.1
263734_at	11	3.3	3	234.4
262324_at	11	3.8	2	278.3
267555_at	11	3.9	2	230.8
254519_at	11	4.0	2	330.7
257935_at	11	4.3	2	265.1
265811_at	11	4.9	3	329.6
264509_at	11	5.0	2	227.8
263390_at	11	5.1	2	890.9
247948_at	11	5.2	2	301.9
262327_at	11	5.3	3	282.7
245506_at	11	5.5	3	272.7
266316_at	11	6.1	2	276.1
249003_at	11	7.0	4	408.5
245348_at	11	7.3	4	295.9
247772_at	11	7.7	3	206.2
259620_s_at	11	8.0	3	252.5
248552_at	11	9.2	3	236.6
255308_at	11	9.4	3	333.8
253399_at	11	10.6	2	343.6
266106_at	11	10.7	2	1202.2
245792_at	11	11.9	3	669.3

Analysis Info   Metrics   Pivot

Fig 2

246745_at, Ymax:4842] [B1798_rep1, P] [M A

246745_at, Ymax:3972] [ATH1_Bc+P_rep1] B

ABSTRACT

Microarray Data Analysis Technique

This invention provides a novel method for the analysis of cross-species microarray
data generated with for example GeneChip® technology. Essentially, oligonucleotide
probes are pre-selected from for example the GeneChip® array by way of cross-
species genomic DNA hybridisation to the GeneChip® array. Probe selection through
database searches with sequences represented on for example GeneChip® arrays is an
alternative. The selected probes are used to construct a library file for gene expression
software designed for the computation of gene expression levels.
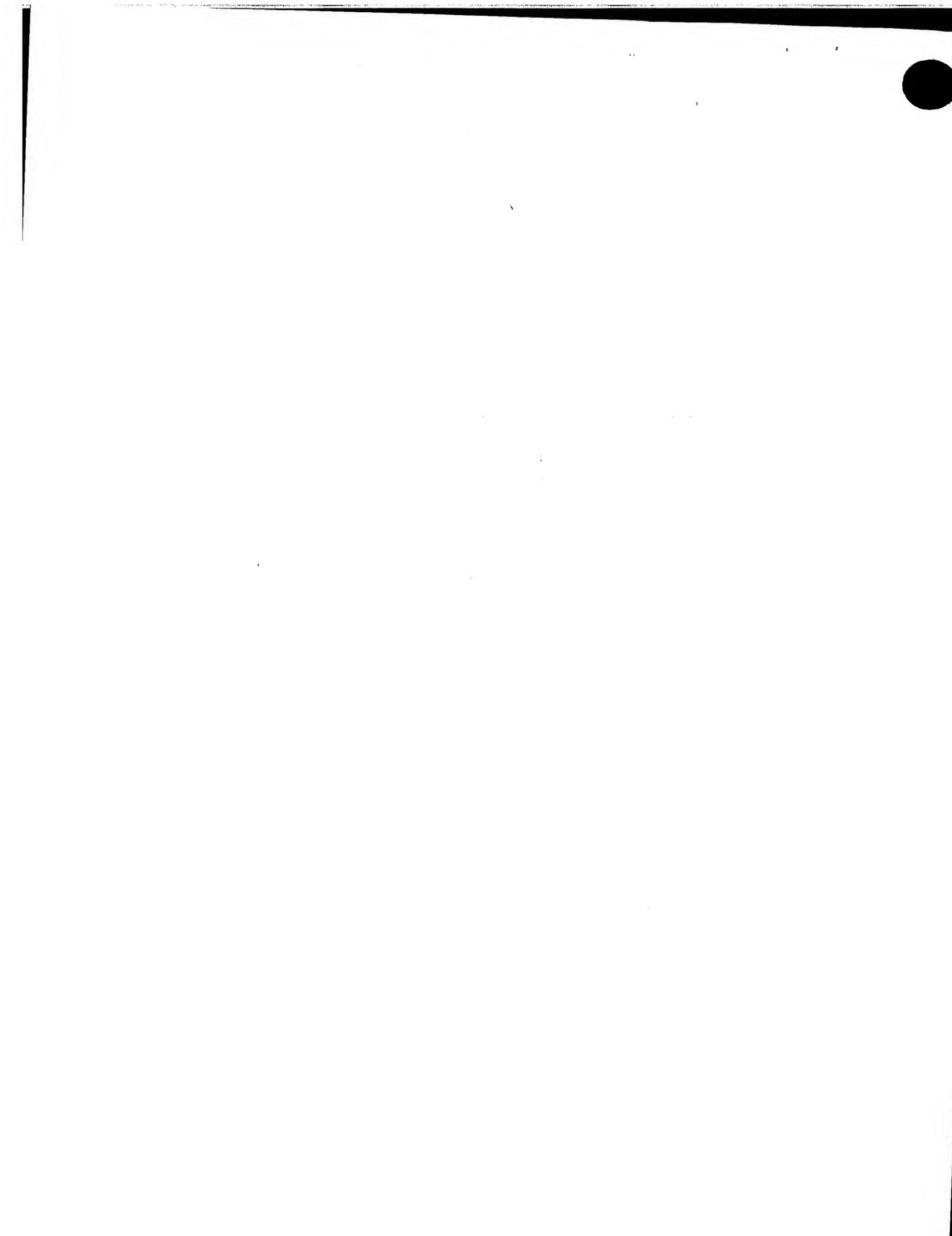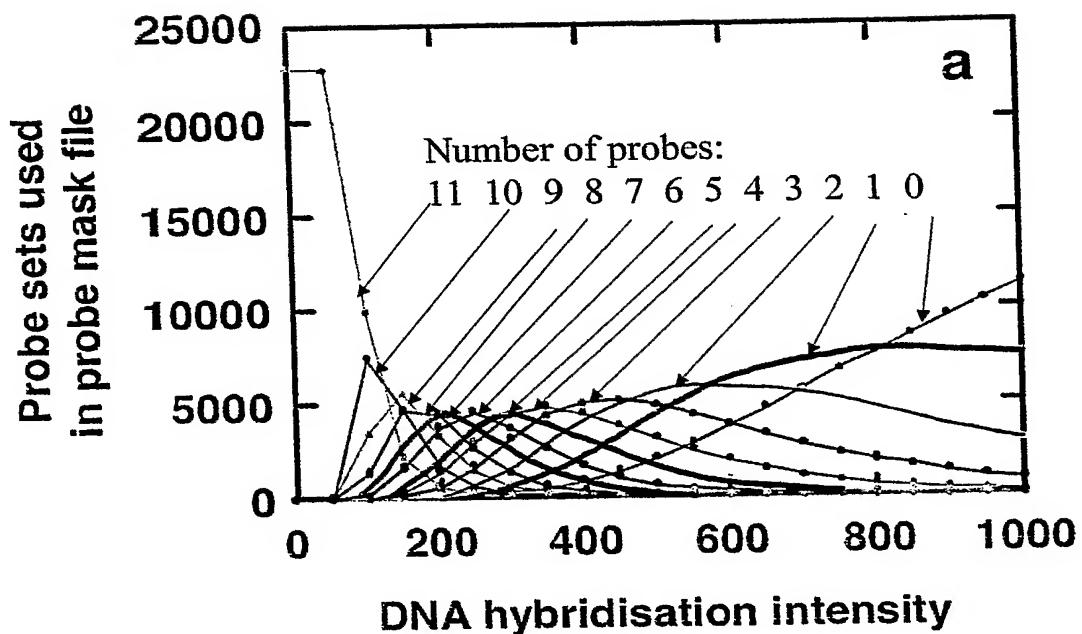
Figure 1a
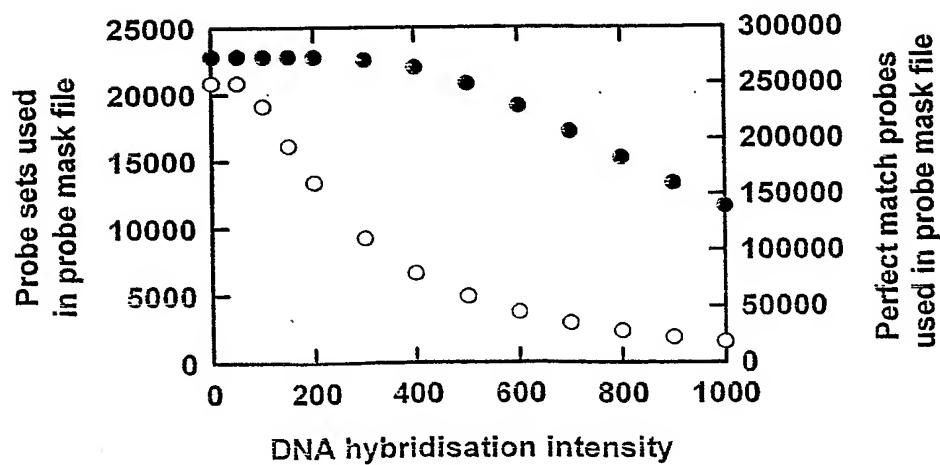
Figure 1b
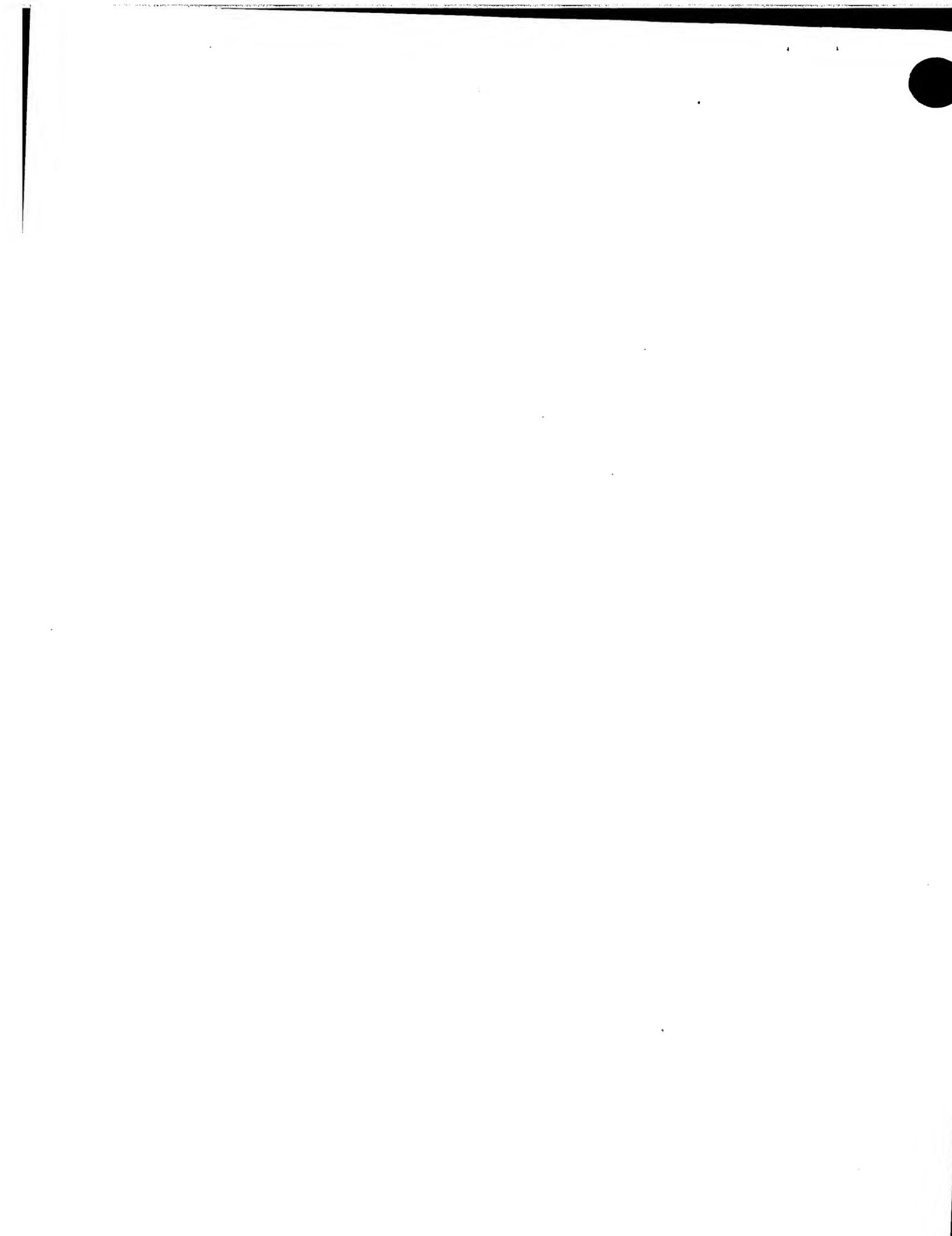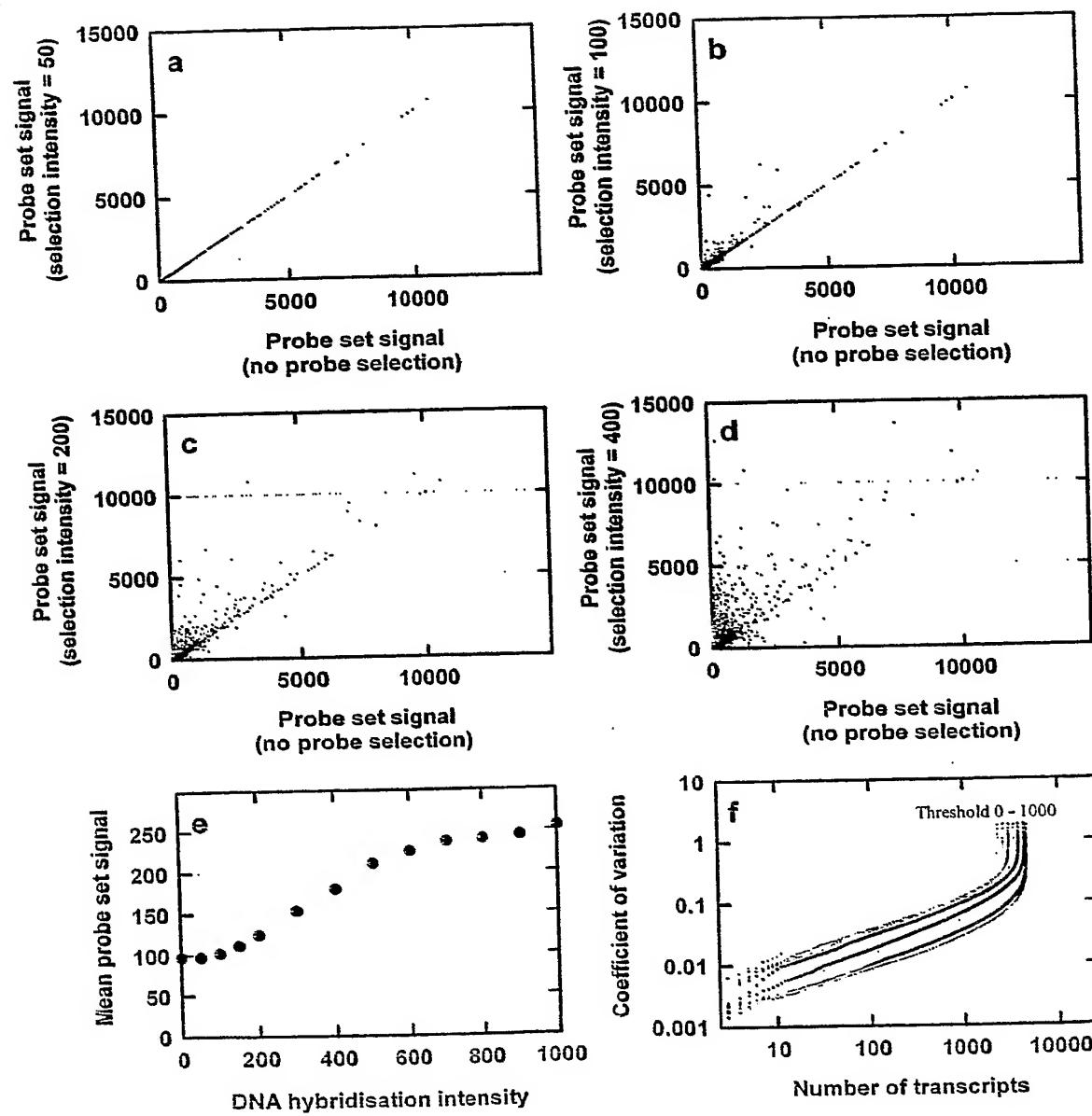
Figure 2
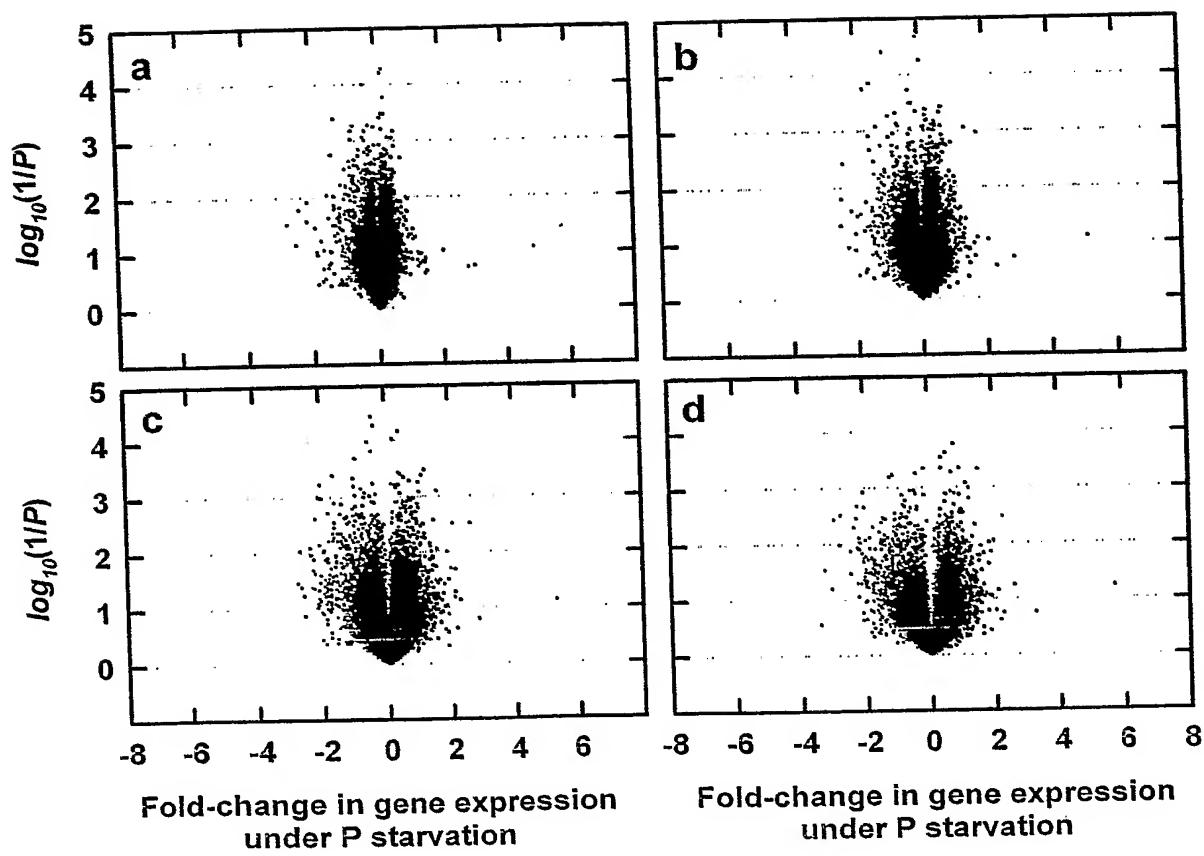
Figure 3
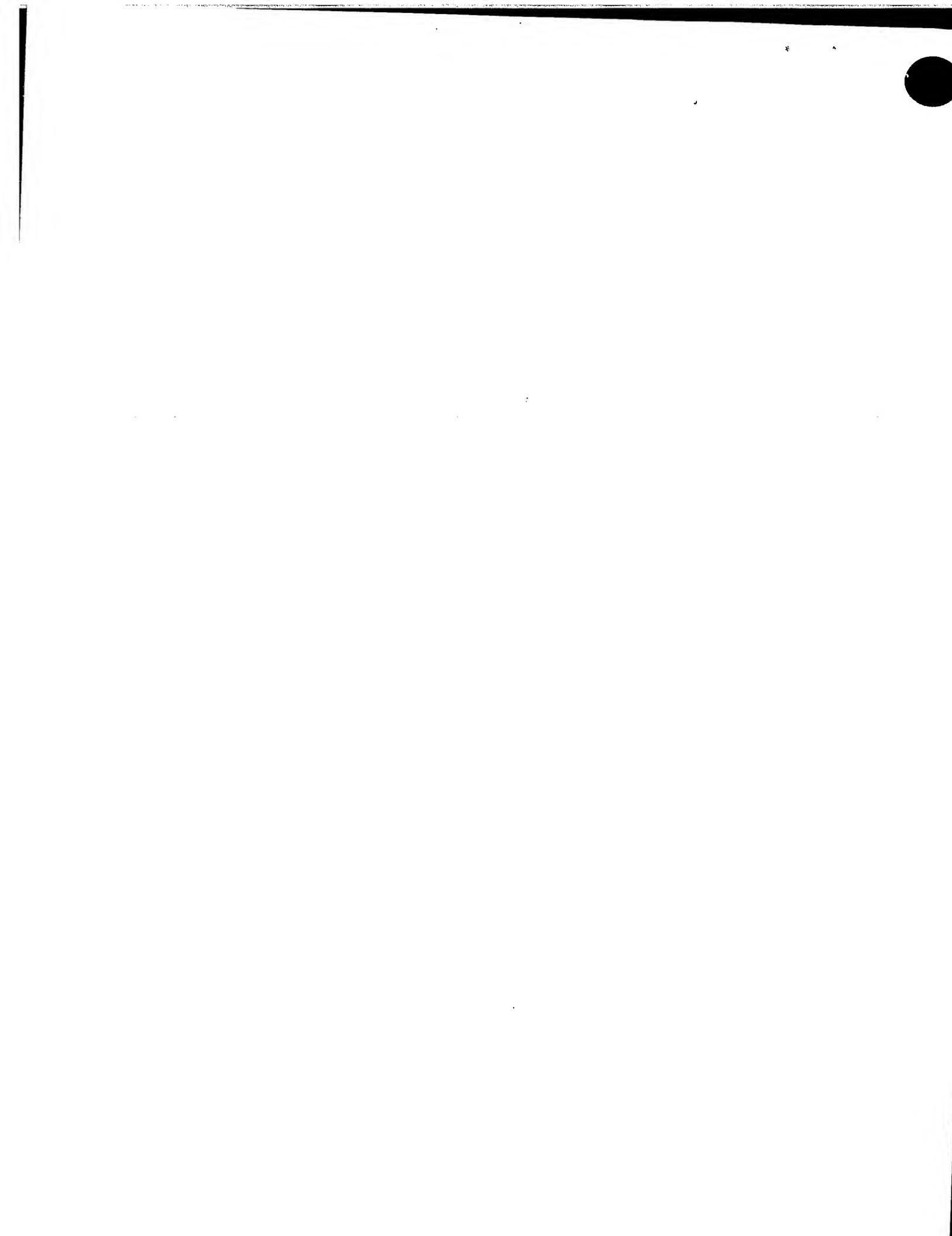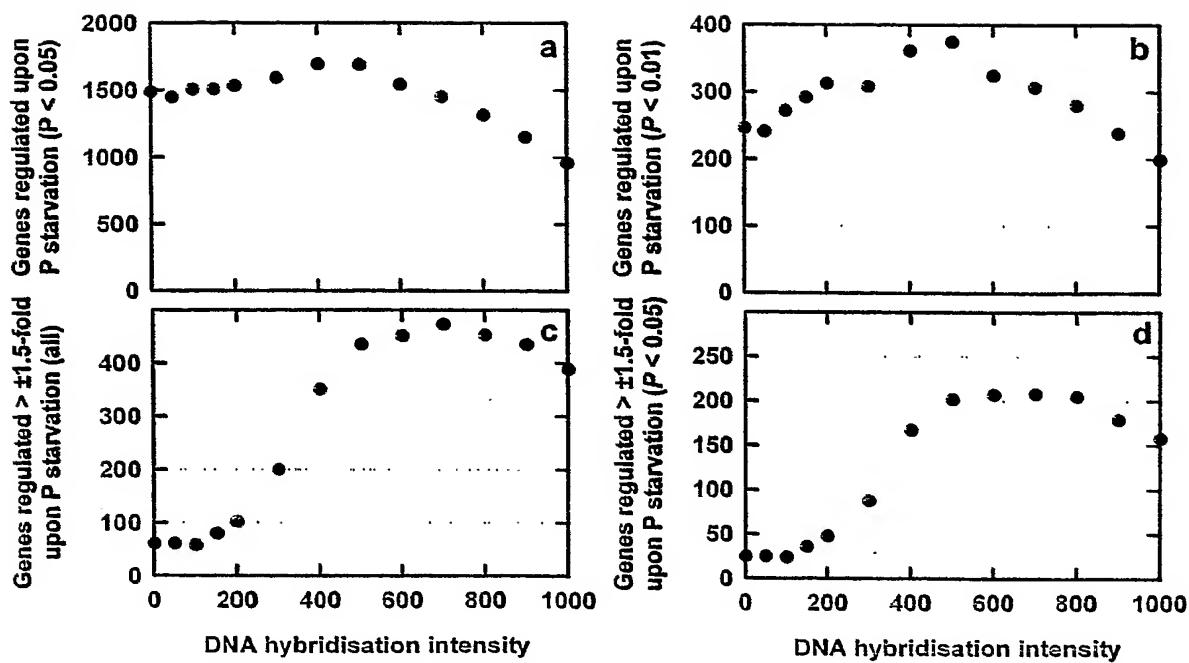
Figure 4
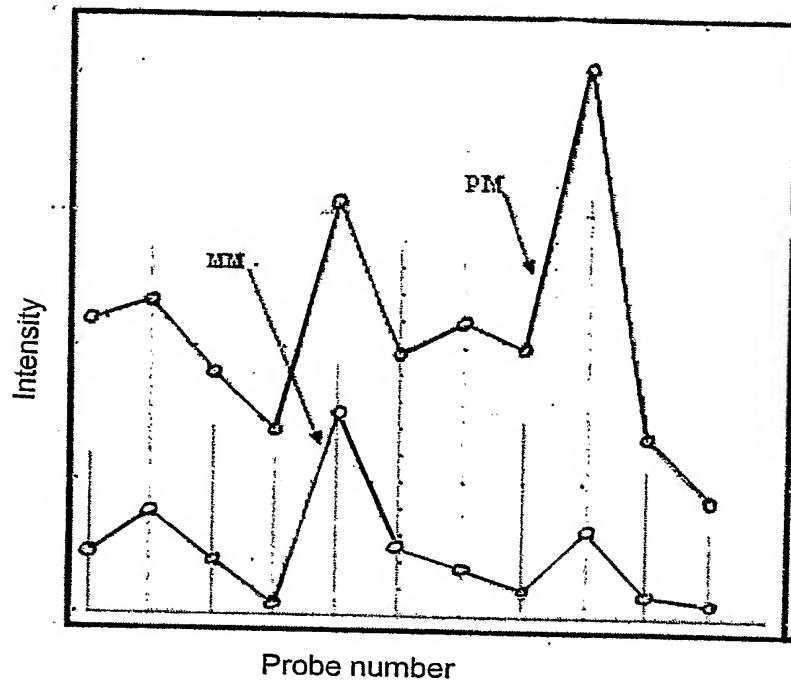
Figure 5a
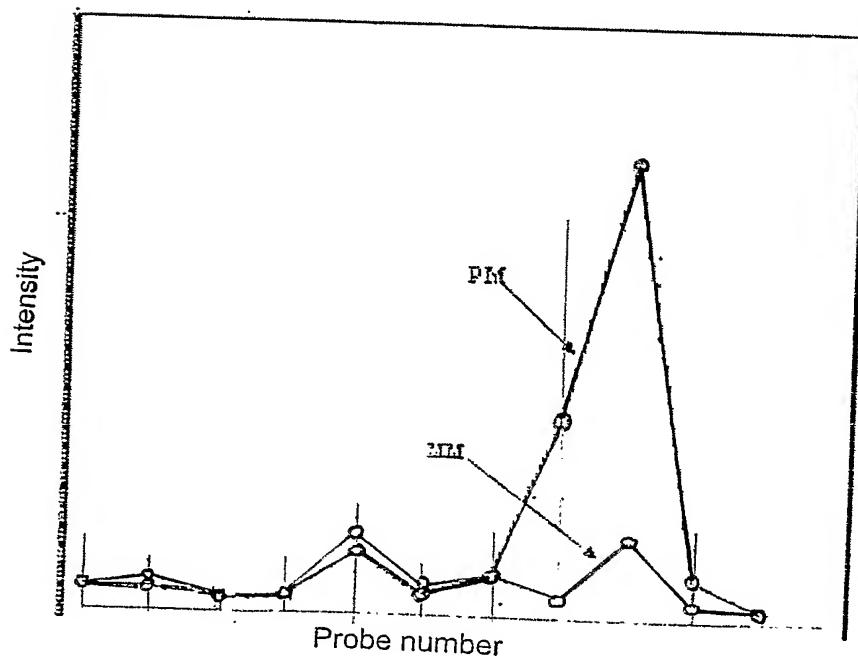
Figure 5b

Figure 6

	ATH1_Bo+P		Brassica_Bo+P	
	Stat Pairs	Signal	Stat Pairs	Signal
250555_at	11	1.4	2	300.5
263634_at	11	1.9	3	1091.4
250286_at	11	2.2	3	299.9
247304_at	11	2.8	4	305.7
259177_at	11	3.1	3	338.1
263794_at	11	3.3	3	234.4
262324_at	11	3.8	2	278.3
267555_at	11	3.9	2	230.8
254519_at	11	4.0	2	330.7
257995_at	11	4.3	2	265.1
265811_at	11	4.9	3	329.6
264509_at	11	5.0	2	227.8
263390_at	11	5.1	2	890.9
247948_at	11	5.2	2	301.9
262327_at	11	5.3	3	282.7
245506_at	11	5.5	3	272.7
266316_at	11	6.1	2	276.1
249003_at	11	7.0	4	408.5
245348_at	11	7.3	4	295.9
247772_at	11	7.7	3	206.2
259620_s_at	11	8.0	3	252.5
248552_at	11	9.2	3	236.6
255308_at	11	9.4	3	333.8
253399_at	11	10.6	2	343.6
266106_at	11	10.7	2	1202.2
245792_at	11	11.9	3	669.3

100

104

120

102

111

114

116

112

106

108

110

Figure 7